

Comparing Standard to Feature-Based Meteorological Model Evaluation Techniques in Bogotá, Colombia

ROBERT NEDBOR-GROSS, BARRON H. HENDERSON, AND JUSTIN R. DAVIS

Engineering School of Sustainable Infrastructure and Environment, University of Florida, Gainesville, Florida

JORGE E. PACHÓN

Centro Lasallista de Investigación y Modelación Ambiental, Universidad de la Salle, Bogotá, Colombia

ALEXANDER RINCÓN, OSCAR J. GUERRERO, AND FREDDY GRAJALES

Secretaría Distrital de Ambiente, Bogotá, Colombia

(Manuscript received 27 January 2016, in final form 20 October 2016)

ABSTRACT

Standard meteorological model performance evaluation (sMPE) can be insufficient in determining “fitness” for air quality modeling. An sMPE compares predictions of meteorological variables with community-based thresholds. Conceptually, these thresholds measure the model’s capability to represent mesoscale features that cause variability in air pollution. A method that instead examines features could provide a better estimate of fitness. This work compares measures of fitness from sMPE analysis with a feature-based MPE (fMPE). Meteorological simulations for Bogotá, Colombia, using the Weather Research and Forecasting (WRF) Model provide an ideal case study that highlights the importance of fMPE. Bogotá is particularly interesting because the complex topography presents challenges for WRF in sMPE. A cluster analysis identified four dominant meteorological features associated with air quality driven by wind patterns. The model predictions are able to pass several sMPE thresholds but show poor performance for wind direction. The base simulation can be improved with alternative surface characterization datasets for terrain, soil classification, and land use. Despite doubling the number of days with acceptable specific humidity, overall acceptability was never more than 10%. By comparison, an fMPE showed that predictions were able to reproduce the air-quality-relevant features on 38.4% of the days. The fMPE is based on features derived from an observational cluster analysis that have clear relationships with air quality, which suggests that reproducing those features will indicate better air quality model performance. An fMPE may be particularly useful for high-resolution modeling (1 km or less) when finescale variability can cause poor sMPE performance even when the general pattern that drives air pollution is well reproduced.

1. Introduction

Meteorological model performance is critical for successful air quality modeling and necessary for regulatory purposes. The standard model performance evaluation (sMPE) techniques for judging model fidelity are based on statistical thresholds developed by the community from various regional studies such as Emery et al. (2001). However, these types of evaluations are likely to show failure for high-resolution modeling in regions of complex topography. Despite sMPE failure, a

meteorological model may be able to successfully pass a feature-based model performance evaluation (fMPE) by reproducing features that are most important for air quality. A region such as Bogotá, Colombia, which has emerging air quality issues, is understudied with low quality surface characterization datasets, and complex topography provides an interesting case study for comparing an sMPE of a meteorological model with an fMPE.

As a result of rapid economic and population growth, air pollution in Bogotá is far above local and international standards. Particulate matter on the order of $\sim 10 \mu\text{m}$ or less (PM_{10}) frequently exceeds Colombia’s daily ($100 \mu\text{g m}^{-3}$) and annual standards ($50 \mu\text{g m}^{-3}$),

Corresponding author e-mail: Robert Nedbor-Gross, rnedbor1@ufl.edu

which are less restrictive than the World Health Organization (WHO) recommended daily ($50 \mu\text{g m}^{-3}$) and annual standards ($20 \mu\text{g m}^{-3}$) (SDA 2013b; WHO 2006). Ozone (O_3) also consistently exceeds the national extremely strict local standard of $80 \mu\text{g m}^{-3}$ (approximately 41 ppb), which will likely be revised soon to be higher. Epidemiological studies have shown significant relationships between both of these pollutants and respiratory/cardiopulmonary diseases and this link has long been known (Dockery et al. 1993). Additionally, studies specific to Bogotá highlight the negative health impacts of air pollution (Arciniegas et al. 2006; Rodríguez et al. 2006; Bell et al. 2011; Pachón et al. 2013). With the aim of ameliorating the situation, Bogotá's environmental agency, the Secretaría Distrital de Ambiente (SDA), has formulated a 10-yr air pollution abatement plan requiring an air quality modeling system to test the efficacy of regulatory strategies (SDA 2010). The first requirement of the air quality modeling system is accurate meteorological predictions.

Bogotá has a unique combination of geography and seasonal conditions. Bogotá is on the western slope of the Andes Mountains at 4.5°N and 74°W at an approximate elevation of 2500 m above mean sea level (MSL), making it the highest metropolitan area with more than 10 million people. Bogotá's seasonal meteorology consists of four alternating periods of rainy and dry weather. The dry seasons are January–March and July–September, and the rainy seasons are April–June and October–December. Winds tend to be strongest in Bogotá in the morning between 0600 and 1100 local time (LT) and in the afternoon between noon and 1800 LT (SDA 2013b).

The complex topography of Bogotá is likely to lead to poor performance in wind speed and direction prediction, which are important components in air quality. In Bogotá, SDA (2013b) showed that diurnal cycles of wind speed and direction are associated with high concentrations of PM_{10} . In regions of complex topography like Bogotá, studies consistently find poor wind direction skill for a standard MPE (Gebhart et al. 2014; Reboredo et al. 2015). Gebhart et al. (2014) used WRF in the Rocky Mountains and concluded that carefully considered physics options did not substantially improve model performance. Reboredo et al. (2015) tested physics configurations with WRF in Bogotá. They modeled eight single days and selected an optimal configuration with marginal differences or improvements from the other parameterizations. All physics parameterizations tested by Reboredo et al. (2015) performed poorly for wind direction. These studies demonstrate the need to focus on other aspects of modeling outside of the

physics configurations, such as surface characterization inputs to improve model performance and demonstrate the inability of meteorological models to pass an sMPE in regions of complex topography.

Previous meteorological modeling studies in other understudied regions like Bogotá suggest sMPE enhancements are possible through improving surface characterization. Understudied regions are generally outside of the United States and Europe, where WRF is most commonly applied, and the default datasets such as the U.S. Geological Survey (USGS) Global 30 arc-s elevation (GTOPO30) topography, land-use, and soil classification data are based on sparse or old observations. However, meteorological models can be improved using alternate inputs for topography (Teixeira et al. 2014), soil categorization (Chang et al. 2014; Cheng and Steenburgh 2005), and land-use classification (Yucel 2006).

There is a clear need for meteorological modeling in Bogotá that addresses the challenges of surface characterization and evaluation within the context of complex topography. This study has three main steps: 1) application of a cluster analysis to characterize the meteorological features in Bogotá that results from the complex topography and their importance for pollutant transport, 2) development of a meteorological simulation using the best available physics options and surface characterization datasets, and 3) comparison of the simulations' ability to pass an sMPE with its ability to reproduce the meteorological features found in the first step using an fMPE.

The cluster analysis reveals that important features for air quality in Bogotá are diurnal local flow patterns. For each cluster, we characterize case studies to illustrate the role of mountains in producing synoptic–thermal interactions that drive local patterns. Initial efforts to simulate local patterns showed poor sMPE, and our results show that improving surface characterization datasets has limited capability for improvement. We propose and implement an alternative fMPE and our results show that simulations are substantially more likely to reproduce critical features than to pass sMPE thresholds. The meteorological features correlate with air pollution regimes, thus suggesting that an sMPE may not correlate with air quality skill.

This manuscript is organized into a methods section (section 2), three results sections (sections 3–5), and our conclusions (section 6). Section 2 describes methods for identifying important meteorological features in the observations with a cluster analysis, WRF surface characterization inputs and physics configurations, the sMPE, and the fMPE. Section 3 details the meteorological features from the observational cluster analysis

and their relation to air quality conditions. [Section 4](#) presents model evaluation results from the sMPE and [section 5](#) discusses the ability of the model to reproduce the important meteorological features with an fMPE. The conclusions are presented in [section 6](#). To our knowledge, there have not been any studies that use a cluster analysis to identify observed meteorological features and test the sensitivity of meteorological models to surface characterization, and, then, compare an sMPE with an fMPE for a region similar to Bogotá.

2. Methods

Meteorological features that are influential for air quality in Bogotá are identified using a k -means cluster analysis. We then simulate the meteorological conditions in Bogotá using WRF (version 3.5.1; [Skamarock et al. 2008](#)) and evaluate the results with an sMPE. Alternative surface characterization input datasets are implemented to assess their impacts on model performance with the goal of improving sMPE. Finally, we use a feature-based fMPE to evaluate WRF's skill at reproducing the features and their frequencies.

a. Identifying observed meteorological features

A cluster analysis can be used to identify features, such as local flow patterns, that are important for air quality ([Davis et al. 1998](#); [Huang and Smith 1999](#); [Thompson et al. 2001](#); [Austin et al. 2015](#)). In this study, a k -means cluster analysis similar to that of [Davis et al. \(1998\)](#) is performed. The k -means method separates data into a number of clusters k on the basis of similarity in selected meteorological parameters.

The k -means clustering algorithm uses as input the number of clusters to generate k and a set of observations to cluster. The algorithm returns a set of k centroids, where all points in one of the k clusters are closest to one of the centroids. Being closest implies that the absolute error of the points in the cluster is minimized with the centroid it is assigned relative to any other centroid. To start, two samples of k centroids are selected from the dataset at random. The first set was seeded for consistency. Then, if they are unequal (i.e., have not converged) one of the samples of k centroids is used to make k clusters. Clusters are made such that each value in the dataset is assigned to a centroid that it is closest to. Then, the centroids are adjusted to be the mean of each respective cluster. The new adjusted centroids are compared with the previous set of centroids for the next iteration, if adjusted centroids have changed, then iterations continue until assignments of clusters and centroids no longer change.

The number of clusters k is chosen based on the “gap statistic” developed by [Tibshirani et al. \(2001\)](#). The statistic compares the “compactness” of clusters with a random distribution of data within the dataset's bounds. The appropriate number of clusters is chosen when the measure of compactness is largest relative to the random distribution's compactness.

Several meteorological variables available from the Red de Monitoreo de Calidad del Aire de Bogotá (RMCAB) observational dataset ([SDA 2013a](#)) were selected for use in the cluster analysis. Among the variables examined were temperature, pressure, mixing height, specific humidity, wind speed, and wind direction at each monitor. Pressure was averaged over all monitors to represent larger-scale processes. The RMCAB dataset has 11 stations that provide hourly data for the variables of interest ([Fig. 1](#)).

By far the most important variables for clustering were morning and afternoon wind speed and direction ([SDA 2013b](#)). All variables were tested but only wind speed and direction were impactful for clustering. Other variables such as temperature, mixing height, solar radiation, specific humidity, and domain-averaged pressure were not impactful for clustering because they do not exhibit nearly as much variability as wind speed and direction in Bogotá. [SDA \(2013b\)](#) reported that winds in Bogotá are only considered to have significant impacts on variability between the hours of 0600–1100 and 1200–1600 LT. Winds at other hours of the day are calm and not considered significant. Therefore, for the cluster analysis, winds were taken as averages of the two intervals of 0600–1100 and 1200–1600 LT.

The k clusters are developed from a year (2012) of observations from RMCAB. The k -means technique requires valid data for all “rows,” where a row contains morning and afternoon wind speed and wind direction for all monitors. Monitors with data unavailable for more than 10% of days in 2012 were excluded. The threshold of 10% was chosen to optimize both the number of stations available for clustering, and an acceptable percent of times that would be included in the cluster analysis. A 10% threshold yields 6 of the 11 stations that meet the threshold ([Fig. 1](#)) on 80% of all possible observation days in 2012 for clustering.

Cluster analysis creates k clusters that have distinct meteorological features, which would ideally correspond to air pollution regimes. Each day and its concentrations of PM_{10} and O_3 are assigned to the clusters to determine which are associated with poor air quality. The air quality conditions associated with each cluster

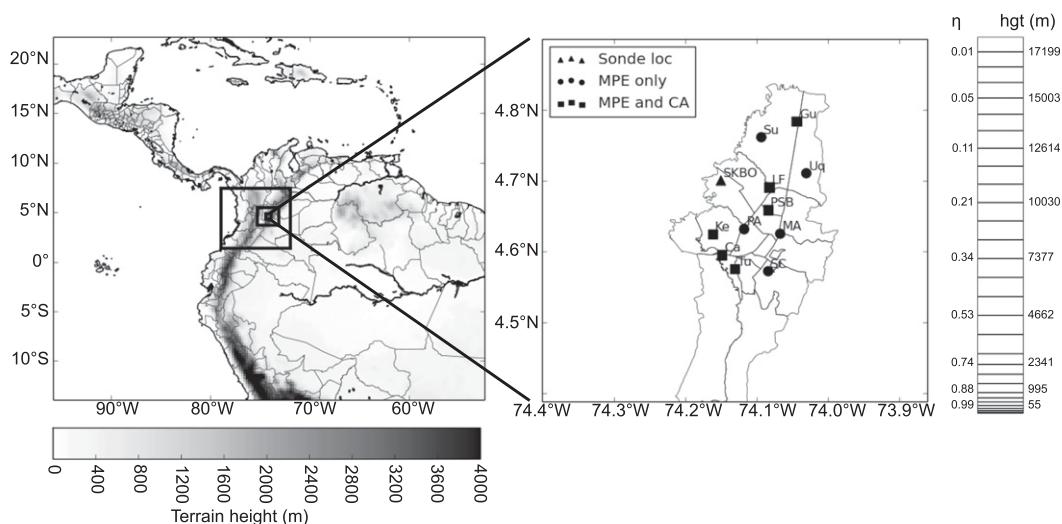


FIG. 1. Domain configuration for WRF simulations. (left) The whole map represents d01, with nested domains shown as boxes and terrain height shaded. (right) Inset showing the d04 domain and the 11 RMCAB stations used for the analysis. Stations represented by squares were used in the cluster analysis (CA) and the MPE. Circles indicate stations used in the MPE only; the triangle shows the location of the sonde data. The vertical structure is on the far right, showing the η levels and their associated heights. Station names and their abbreviations are Guayamaral, Gu; Suba, Su; Usaquen, Uq; Las Ferias, LF; Parque Simon Bolivar, PSB; Puente Arande, PA; Kennedy, Ke; MAVDT, MA; Cazuca, Ca; Tunal, Tu; San Cristobal, SC; and the sonde location, SKBO.

are identified as distinctly different using a Mann–Whitney rank sum test (Mann and Whitney 1947).

b. Model domain configuration and simulation periods

WRF was run for Bogotá using nested domain simulations with boundary and initial conditions from the National Centers for Environmental Prediction Final Operational Global Analysis (FNL; NCEP 2000). The domains (Fig. 1) are centered over Bogotá and use a 3-to-1 parent-to-child nesting ratio. The outermost domain (d01) has a 27-km resolution and encompasses all of Colombia, much of the surrounding region, and the Caribbean Sea while using a spectral nudging approach. The first nested domain (d02), with 9-km resolution, represents much of Colombia while minimizing boundary placement on complex terrain. The second nested domain (d03), with a 3-km resolution, encompasses the Colombian department (similar to a U.S. state) of Cundinamarca. The innermost domain (d04) has a 1-km resolution, is 64 km \times 64 km and is centered over Bogotá. Vertically, the simulations have 30 layers for all domains with the height of the first layer being 55 m. The vertical structure is also included in Fig. 1. Data were output from WRF hourly for all domains.

We simulate two 25-day periods in 2012 when high-pollution episodes occurred for a dry season and a rainy season (50 total simulation days). The periods spanned 7 February–3 March and 30 September–24 October and

are referred to hereafter as the February and October periods. Each 25-day period was broken up into five overlapping 5.5-day segments. The first 12 h of each segment (prior to the start date at local time) were discarded as spinup.

c. Physics configurations

Multiple physics configurations were evaluated for their influences on model performance and two were selected for evaluation with improved surface inputs. A physics configuration is a combination of WRF's physics schemes for radiation, cumulus convection, microphysics, the planetary boundary layer (PBL), and land surface model (LSM) (full details can be seen in appendix A). Two of the configurations were chosen for the sMPE and evaluation of surface characterization inputs based on the PBL schemes. PBL schemes are significant because of their influence on humidity. The first configuration chosen is based on the default WRF configuration (DF), and the second configuration is based on Nielsen-Gammon (2001), which was used in the Houston–Galveston, Texas, region (NG). The DF configuration uses the Yonsei University (YSU) PBL scheme, which has a known improvement from the MRF PBL scheme used by the NG configuration. The YSU PBL scheme represents the PBL height better than does the MRF PBL scheme used by the NG configuration because of the explicit treatment of entrainment (Hong et al. 2006; Cohen et al. 2015).

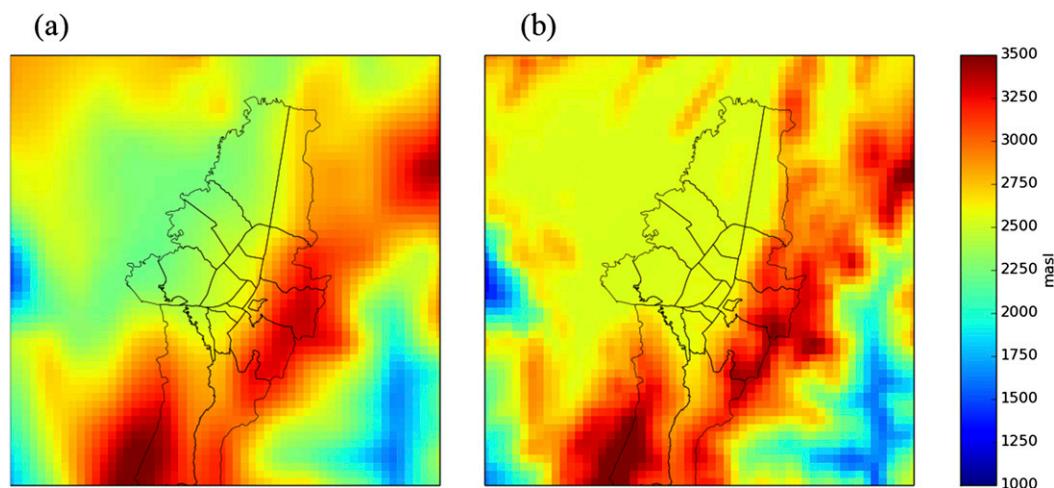


FIG. 2. Spatial plots showing terrain (m MSL) for domain d04 using (a) USGS GTOPO30 and (b) ASTER data.

d. Surface characterization inputs for WRF

The default surface characterization datasets are highly accurate for use in the United States but may return subpar results outside of the country. The WRF Preprocessing System (WPS) allows the user to select surface characterization datasets (e.g., topography, land use, and soil classification) to be used by WRF. WPS then prepares data for each domain. We tested three surface characterization datasets that are alternatives to the standards for influence on model skill: topography, soil categorization, and land use. These datasets are tested with the two physics configurations: DF and NG. We also test the propagation of soil temperature and moisture from each 5-day segment to segment. For each surface characteristic, we demonstrate the difference between the standard WRF dataset (at the highest available resolution) and an alternative data source described further below.

1) TOPOGRAPHY

Bogotá has complex topographic features, which need to be sufficiently resolved in order to yield optimal meteorology. The standard USGS GTOPO30 topography interpolated by WPS to the selected domain has deficiencies for Bogotá. Figure 2a shows that the GTOPO30 topography has a local depression in the northwest portion of the d04 domain. Local researchers (Rincon and Rojas 2014, 2015) note that the actual terrain is much flatter in this region, and the depression is likely the result of interpolation from sparse data. Alternatively, we use the Advance Spaceborne Thermal Emission and Reflection Radiometer (ASTER) topographic dataset (Abrams 2000).

ASTER has a 30-m resolution and has been more accurate than the GTOPO30 in select regions outside the United States (Nikolakopoulos et al. 2004). To implement ASTER for WRF, we converted original GeoTIFF-formatted ASTER files to the necessary Band Interleaved by Line (.bil) format for compatibility with WPS. Figure 2b shows that topography derived from the ASTER dataset does not have the artificial depression in the northwest and has a higher mean height and stronger spatial gradients. For ASTER simulations, ASTER replaces GTOPO30 for all domains except d01, where ASTER's resolution does not lead to significant changes.

2) LAND USE

Land use has been impactful for atmospheric state variables including wind, temperature, and humidity (Yucel 2006). An image of Bogotá shown in Fig. 3a, generated using Google Earth, displays the urban area of the city with a generally gray color, with surrounding vegetation. The default USGS land-use dataset in Bogotá (Fig. 3b) does not cover enough of the urban area with the urban category and poorly represents the city's surrounding vegetation. We tested simulations using WRF's alternative available land-use dataset, the Moderate Resolution Imaging Spectroradiometer (MODIS) (Boston University 2012), as seen in Fig. 3c, which has a higher resolution and can be more accurate outside of the United States (Cheng and Steenburgh 2005). MODIS provides a better spatial representation of the urban area of Bogotá and the forested regions surrounding the city. These two datasets are juxtaposed in Figs. 3b and 3c with a joint color scale representative of both by combining similar categories.

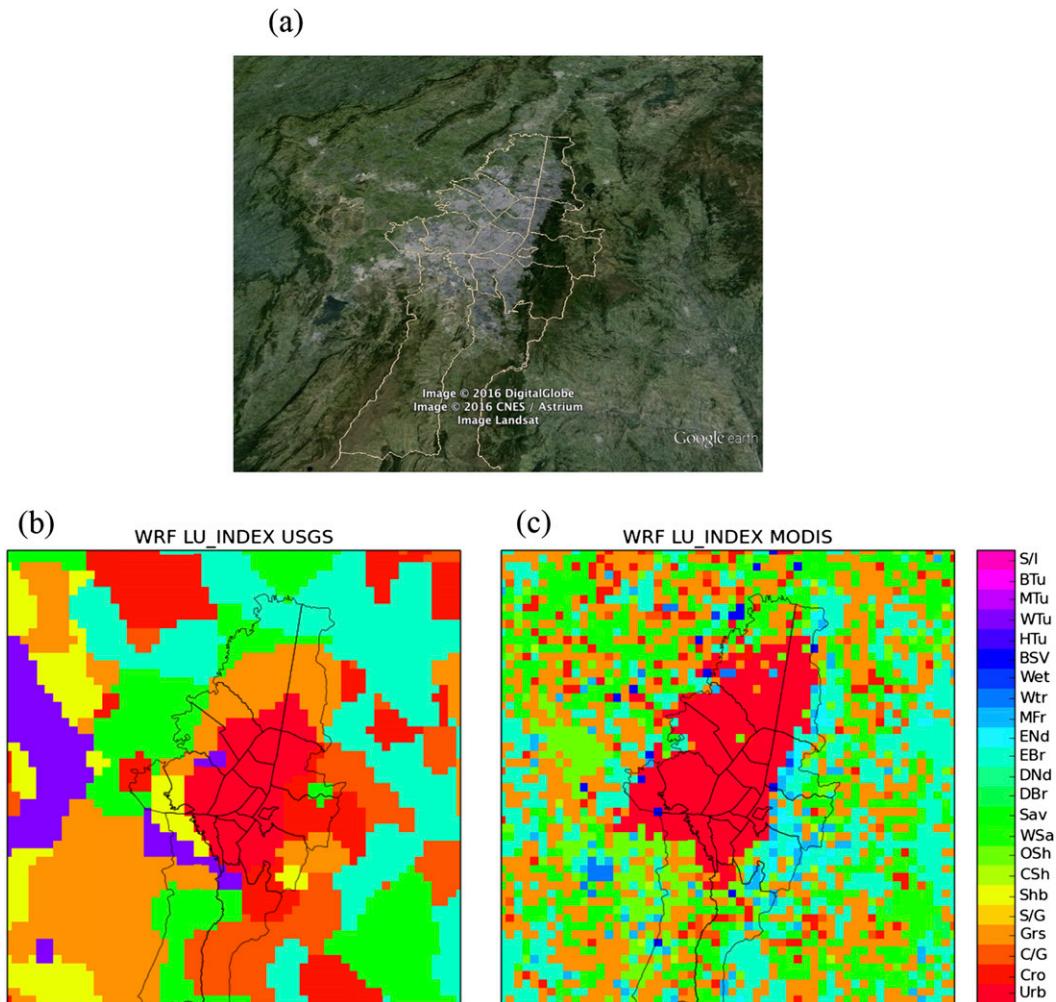


FIG. 3. (a) Google Earth image of Bogotá and the surrounding area. Also shown in juxtaposition are the land-use categories in the d04 domain derived from the (b) USGS and (c) MODIS datasets. Categories were mapped onto a single bar representative of each dataset (snow/ice, S/I; barren tundra, Btu; mixed tundra, MTu; wooded tundra, WTu; herbaceous tundra, HTu; barren/sparse vegetation, BSV; wetland, Wet; water, Wtr; mixed forest, MFr; evergreen needleleaf, ENd; evergreen broadleaf, EBr; deciduous needleleaf, DNd; deciduous broadleaf, DBr; savanna, Sav; woody savanna, WSa; open shrubland, OSh; closed shrubland, CSh; shrubland, Shb; shrub/grassland, S/G; grassland, Gr; crop/grassland, C/G; cropland, Cro; and urban, Urb).

3) SOIL CLASSIFICATION

The standard USGS soil classification dataset for WRF is poorly resolved in Bogotá. Figure 4a shows that the USGS has a distinctly blocky pattern and only two categories. As an alternative, we used the Harmonized World Soil Database (HWSD) with 1-km resolution (Batjes 2009). HWSD has soil data in percentages of soil type. Since WRF requires soil classification data in USGS classifications, we convert the HWSD percentages of clay, silt, and sand to the proper classifications in USGS format. Figure 4b shows that HWSD has a moderately better representation of d04 with four classifications and a geographic pattern that more closely

resembles the mountainous east, central plateau, and agricultural west.

4) SOIL STATE PROPAGATION

Soil classification and land use indirectly affect meteorological modeling skill through soil moisture and temperature, but the cumulative effect that takes time to develop requires special treatment. The cumulative effects on soil moisture and temperature are typically lost during each 5-day segment reinitialization. This may be important for HWSD simulations, which have a different soil class dataset than NCEP FNL. As an alternative to reinitialization and similar to a previous study over

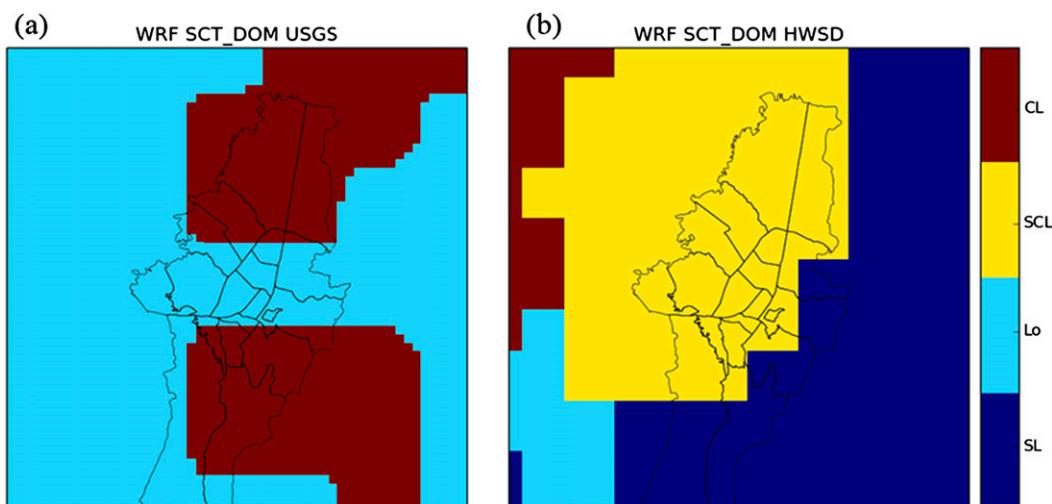


FIG. 4. Soil category data for domain d04 from (a) USGS and (b) HWSD (sandy loam, SL; loam, Lo; sandy clay loam, SCL; clay loam, CL).

the United States (Gilliam and Pleim 2010), we try propagating soil moisture and temperature data from the end of a 5-day segment to the start of the following segment’s spinup period to allow for cumulative deviation from the NCEP (2000) initial conditions. The initial conditions for WRF are shown in Fig. 5a for the third segment in February following the spinup, and by contrast the propagated soil moisture for the same time is shown in Fig. 5b using the HWSD soil classification. Using propagation shows a clear difference in moisture particularly over the mountainous east of d04, likely accounting for more runoff over time.

5) SURFACE CHARACTERIZATION SUMMARY

With all combinations of terrain, soil classification, soil moisture propagation, and land use, there are a total

of 16 characterizations, as seen in Table 1. There are eight simulations for each standard surface characterization and eight simulations for each of the alternative datasets.

e. Standard model performance evaluation

The sMPE can be used to determine if the model is suitable for regulatory air quality modeling and to understand the importance of applying improved surface characterization datasets. However, an sMPE will not be able to test if WRF can recreate the features important for air quality as determined by the cluster analysis. Each model simulation is assessed using benchmarks for “acceptability” in air quality applications. “Acceptability” is characterized based on a synthesis of the broader community’s mesoscale sMPE. Tesche et al.

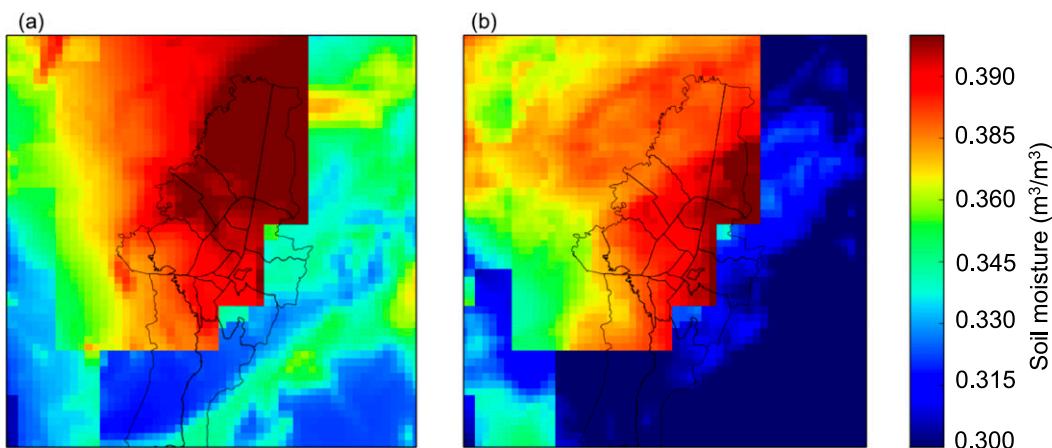


FIG. 5. Soil moisture from WRF input for the DF configuration segment 3 during February (a) without propagation (GHUR) and (b) with propagation (GHUP), both using the HWSD soil classification.

TABLE 1. Input configurations from varying terrain, land-use, and soil category data, as well as soil propagation.

Simulation	Terrain	Soil category	Land use	Soil propagation
GUUR	GTOPO30	USGS	USGS	No
GUUP	GTOPO30	USGS	USGS	Yes
GUMR	GTOPO30	USGS	MODIS	No
GUMP	GTOPO30	USGS	MODIS	Yes
GHUR	GTOPO30	HWSD	USGS	No
GHUP	GTOPO30	HWSD	USGS	Yes
GHMR	GTOPO30	HWSD	MODIS	No
GHMP	GTOPO30	HWSD	MODIS	Yes
AUUR	ASTER	USGS	USGS	No
AUUP	ASTER	USGS	USGS	Yes
AUMR	ASTER	USGS	MODIS	No
AUMP	ASTER	USGS	MODIS	Yes
AHUR	ASTER	HWSD	USGS	No
AHUP	ASTER	HWSD	USGS	Yes
AHMR	ASTER	HWSD	MODIS	No
AHMP	ASTER	HWSD	MODIS	Yes

(2001) compiled MPEs from mesoscale meteorological simulations used for regulatory applications. Emery et al. (2001) added recommendations about benchmarks for statistics applied to individual days. Because of Bogotá's complex terrain, we combine the recommended evaluation thresholds with those from McNally (2009). We use the proposed daily benchmarks, shown in Table 2, for the simulations.

Model performance statistics are compared with benchmarks for four important variables: wind speed at 10 m above ground (WS10), wind direction at 10 m (WD10), 2-m temperature (T2), and 2-m specific humidity (Q2). Model variables are output in WRF as specified in Skamarock et al. (2008). For each variable, a combination of root-mean-square error (RMSE), mean bias (MB), mean error (ME), and index of agreement (IOA) are used. Wind direction is only evaluated when observational wind speeds are greater than 0.5 m s^{-1} . We examine the number of "acceptable" days, where a day is considered acceptable if all metrics for a variable meet the benchmark on that day. Furthermore, we evaluate joint acceptability, which occurs when all variables' metrics meet their respective benchmarks on a given day. Since WD10 error is a highly difficult metric, we also consider "joint acceptability" for all variables without it. Not including WD10 error places strong emphasis on WD10 bias, which is the more important wind direction metric for air quality because it indicates whether pollutants will arrive at a monitor as opposed to when they will arrive. For regulatory standards, air quality variables are typically averaged over several hours and/or daily maxima are used, placing further emphasis on the importance of a low bias for WD10.

TABLE 2. Benchmarks for meteorological performance based on Emery et al. (2001) and McNally (2009).

Variable	Metric	Benchmark
Wind speed	RMSE	$\leq 2.5 \text{ m s}^{-1}$
	Absolute value of MB	$\leq 1.5 \text{ m s}^{-1}$
	IOA	≥ 0.6
Wind direction	Mean gross error	$\leq 55^\circ$
	Absolute value of MB	$\leq 10^\circ$
	IOA	≥ 0.8
Temperature	Mean gross error	$\leq 3.5 \text{ K}$
	Absolute value of MB	$\leq 2 \text{ K}$
	IOA	≥ 0.8
Specific humidity	Mean gross error	$\leq 2 \text{ g kg}^{-1}$
	Absolute value of MB	$\leq 1 \text{ g kg}^{-1}$
	IOA	≥ 0.6

The simulations were evaluated against local observational data from RMCAB for the two innermost domains and against Meteorological Assimilation Data Ingest System (MADIS; NCEP 2014) stations for the outer two domains. Three stations had specific variables excluded from the MPE based on analysis from the RMCAB quality assurance team. The San Cristobal (SC) station's wind data were excluded for the February period and the Ministerio de Ambiente, Vivien da y Desarrollo Territorial (MAVDT) station's wind direction was excluded for October because of unreliable data. Also, we excluded Guaymaral's specific humidity as a variable because, although this site is located in a rural region of Bogotá, it is within 30 m of a large paved area, which, according to Bailey (2000), is inappropriate for specific humidity. Vertical performance is also used to test WRF's ability to accurately resolve mesoscale vertical structure. Comparisons were made between WRF and radiosonde ("sonde") data collected at the El Dorado International Airport (SKBO) daily at 1200 UTC.

f. Feature-based model performance evaluation

Cluster analyses can be used to help identify important features in modeling studies, and evaluations based on these features can show similar levels of performance to threshold-based analyses (Dormann et al. 2013). After using a cluster analysis to identify the important meteorological features from the observational dataset (see section 2a) that are predictors of air quality, we determine if the model is able to reproduce those features. This feature-based analysis, fMPE, is potentially more viable than an sMPE, particularly for new modeling locations. At such locations, surface datasets may not be very accurate, and if there is no immediate opportunity for improvement, it is likely that a simulation will fail an sMPE. Instead, an fMPE is more likely to demonstrate that a model is applicable for air quality modeling. The model's ability to reproduce the features

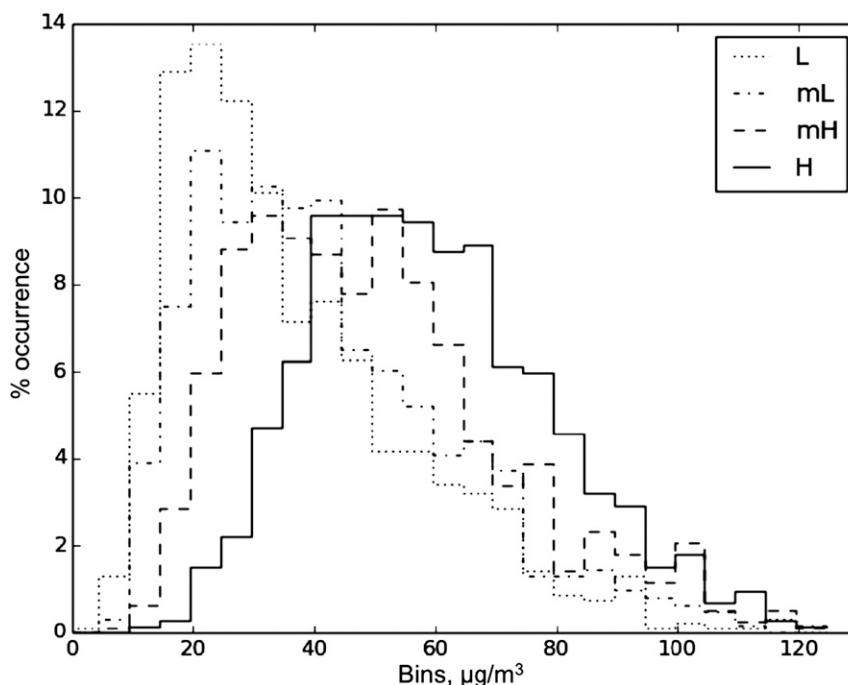


FIG. 6. The PDF of the daily 24-h average PM_{10} concentration from observations for each of the four clusters. Each bin's width is $5 \mu\text{g m}^{-3}$.

is evaluated based on the accuracy of simulated day assignments of the clusters.

Prediction skill is nominally tested by rejecting accuracy as random chance. The null hypothesis is that every day, each cluster has an equal chance of occurrence, making the total number of occurrences of each of the k clusters in a period equally likely. The assumption is that model skill is not driving cluster assignment. The null hypothesis is evaluated using a Bernoulli test (binomial test) of overall assignment accuracy (rejected if $p < 0.05$). If the null hypothesis is rejected, the model nominally reproduces the accuracy of the clusters and, therefore, the air quality tendency.

3. Observational cluster analysis

First, the cluster analysis is used to reveal the meteorological features, such as diurnal wind patterns, that lead to certain air quality conditions. The “gap statistic” (described in section 2a) helped determine that the optimal number of clusters to be considered, k , should be 4. While the clustering method was not based on air quality, the flow patterns associated with clusters did show a relation to air quality conditions, particularly PM_{10} . The clusters' probability density functions (PDFs) showing the distributions of PM_{10} concentrations can be seen in Fig. 6. The clusters will hereafter be referred to

as the low cluster (L), the midlow cluster (mL), the midhigh cluster (mH), and the high cluster (H), each with an average daily PM_{10} concentration from observations of 37, 43, 50, and $61 \mu\text{g m}^{-3}$, respectively. The O_3 concentrations are similar across clusters and low because of strong NO_x titration, which leads to consistently low O_3 in the region. A Mann–Whitney rank sum test was used to determine whether the air quality conditions associated with each cluster were distinct. The test reports a p value for each cluster's relation to all other clusters that needs to be less than 0.05. The maximum p value was 0.02, demonstrating that unique air quality conditions are associated with each cluster.

Figure 7 shows the combination of morning and afternoon wind patterns for each cluster along with associated PM_{10} and O_3 concentrations. Arrows outlined in black were included in the cluster analysis and arrows not outlined were not included in the clustering because of intermittent data availability such that the minimum data availability requirement (see section 2a) was not met. The arrows from intermittently available stations are shown even though they were not used in the cluster analysis. The low PM_{10} cluster L shows strong morning and afternoon winds flowing generally toward the northwest and away from the mountains, likely driven by synoptic winds. The mL cluster has weaker winds that flow either west or north, both directions being away

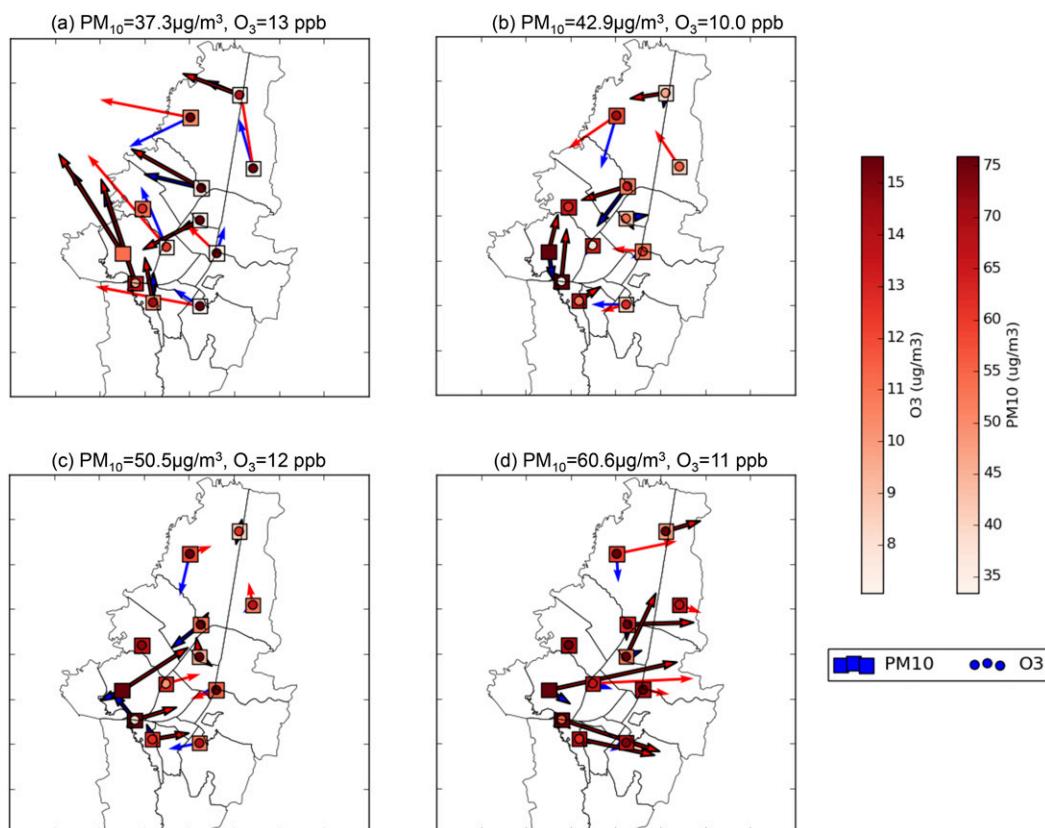


FIG. 7. Wind flow patterns from observations associated with each cluster for morning (blue) and afternoon (red) winds at the monitor location. Arrows that are outlined in black were included in the analysis; arrows not outlined were filtered out. Shown are the (a) L, (b) mL, (c) mH, and (d) H clusters. The arrows represent the direction and distance an air parcel would travel in 1 h. The squares and circles at the monitor locations show PM₁₀ and O₃ concentrations, respectively. There is one extra station plotted without arrows that only reported air quality data.

from the mountains and likely not impacted by synoptic-scale flow. The mH cluster shows several stations with winds flowing upslope toward the mountains in the east during the afternoon and downslope and away from the mountains during the morning, indicating a thermally driven pattern. The H cluster's winds are generally strong, indicating synoptic flow, and would transport pollutants directly toward the mountains in the east.

The concentrations associated with the wind patterns suggest that the mountains in the region play a significant role in pollutant entrapment similar to other studies in regions of complex topography with mountains on one side of the region (Lu and Turco 1995; Yao et al. 2014). The L cluster has strong winds flowing downslope; therefore, pollutants are likely flushed out of the region, leading to lower concentrations. When winds are flowing toward the mountains, as in the H cluster, there is likely to be transport blocking along with air parcels vertically oscillating on the upstream side of the mountain, which would prevent long-range transport and

likely lead to the higher concentrations (Yao et al. 2014). This vertical oscillation can occur because the mountain range is sufficiently long and perpendicular to the flow.

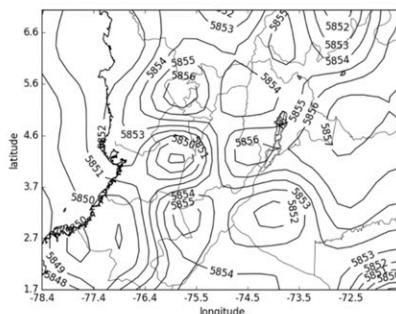
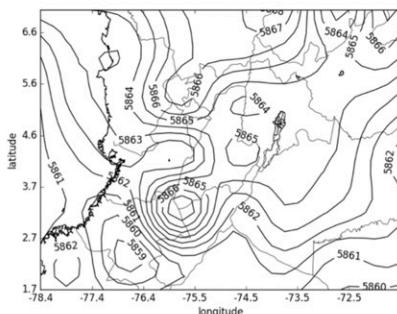
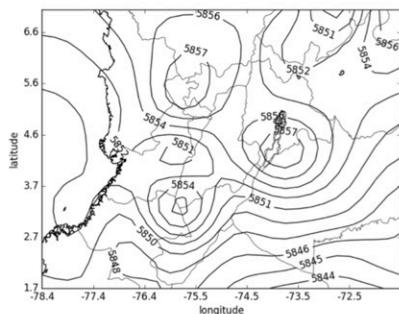
The L and H wind patterns occur when synoptic wind patterns are dominant, while the mH and mL clusters tend to occur when winds are either thermally driven or weak and synoptic flow is not a major factor. When there is strong synoptic flow, it is typical that pollutant concentrations are flushed out and tend to be lower. This is true for the L cluster when winds are easterly, but when the winds are westerly, for the H cluster, the highest concentrations tend to occur. This is indicative of the pollutant trapping. When synoptic flow is not a factor, winds are thermally driven and there tends to be morning winds downslope from the mountains and afternoon winds upslope (SDA 2013b), as seen for the mH cluster.

To demonstrate these effects on pollutant transport, we present a case study of an H day (17 February 2012), an mH day (11 October 2012), and an L day (2 March

(a) Feb. 17th GHT (H)

(b) Oct. 11th GHT (mH)

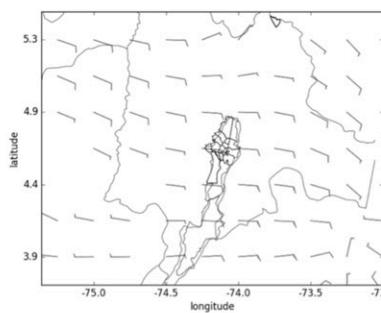
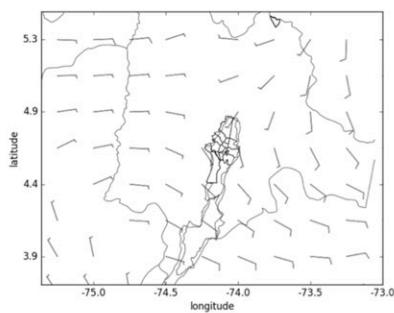
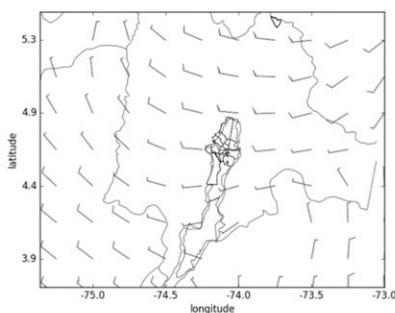
(c) Mar. 2nd GHT (L)



(d) Feb. 17th Wind (H)

(e) Oct. 11th Wind (mH)

(f) Mar. 2nd Wind (L)



(g) Feb. 17th Diurnal (H)

(h) Oct. 11th Diurnal (mH)

(i) Mar. 2nd Diurnal (L)

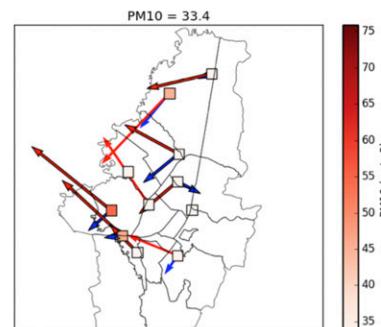
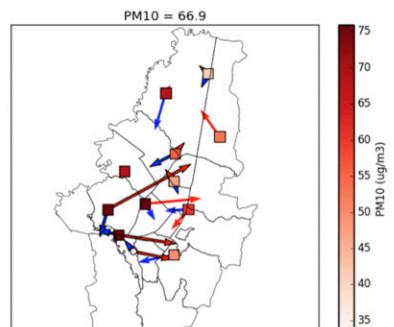
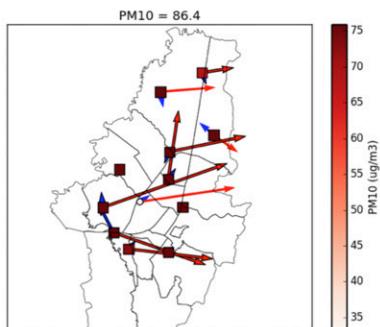


FIG. 8. Case study showing the 500-hPa GHT for (a) H, (b) mH, and (c) L days for domain d02 from gridded NCEP FNL data averaged between 0700 and 1300 LT. (d)–(f) As in (a)–(c), but showing the first layer winds around Bogotá for the d03 domain for the same time averaging from NCEP FNL; these are shown as wind barbs with a half bar as 2.5 kt (where 1 kt = 0.51 m s⁻¹) and a full bar as 5 kt. (g)–(i) As in (a)–(c), but showing the morning (blue) and afternoon (red) winds and PM₁₀ concentrations.

2012). Figures 8a–c shows the 500-hPa geopotential height (GHT) from the NCEP FNL dataset gridded onto d02 and averaged between 0700 and 1300 LT, and the diurnal winds in Figs. 8g–i. The H day has a high

just south of Bogotá along with synoptic winds flowing from the west. The diurnal pattern shows weak morning winds, with strong afternoon winds flowing toward the mountains. The PM₁₀ concentrations on this day remain high at an average of 86.4 μg m⁻³ indicating

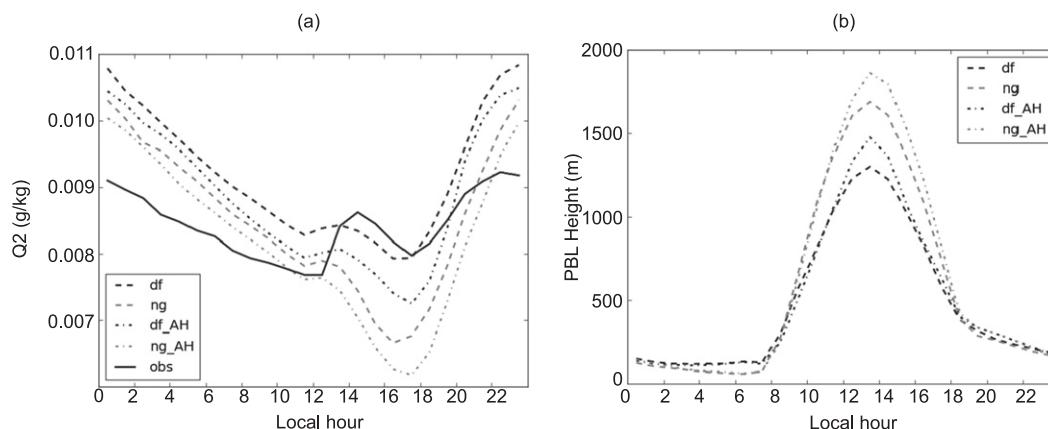


FIG. 9. Daily average (a) specific humidity and (b) PBL heights averaged over all monitor locations for the combined periods in local time.

pollutants did not exit the region and were trapped. On the mH day, there is not a significant system over Bogotá, and surrounding winds are weak and scattered, leading to thermally driven local winds. The mH day winds are downslope in the morning and upslope in the afternoon. The average PM_{10} concentration for the mH day was $66.9 \mu\text{g m}^{-3}$, which is still high because of the afternoon westerly winds but not as high as the H day. For the L day, synoptic winds are flowing from the east and lead to diurnal local winds flowing generally easterly. The average concentration of PM_{10} is $33.4 \mu\text{g m}^{-3}$.

Based on the case study and conditions from the cluster analysis, it can be concluded that in a region such as Bogotá, with mountains that are tallest on one side, wind speed and wind direction are dominant factors determining air quality conditions.

4. Standard model performance evaluation results

The model was run for 35 total simulations (five physics configurations; 16 inputs for two physics configurations). The full evaluation for all configurations is presented in [appendix B](#) for the February and October periods with the two configurations most influenced by surface characterization changes analyzed here. Outer domain analysis (for d01, d02, and d03) is also shown in [appendix C](#) and was similar for all simulations; therefore, this study focuses on performance in d04.

a. Standard model performance evaluation of physics options

The DF and NG physics configurations had common levels of performances for most sMPE criteria, with Q2 as a notable exception. Both configurations performed

very well for T2, moderately for WS10 and Q2, and very poorly for WD10. Vertically, model performance is weakest and most variable near the surface, and stronger and more consistent aloft. A full evaluation (see [appendix B](#)), including vertical analysis (see [appendix D](#)), is presented later.

For WS10, WRF performs fairly well for the DF and NG configurations, with 29 and 28 acceptable performance days out of 50. The model performed poorly for WD10, particularly ME. The DF and NG configurations only have between 4 and 3 days accepted out of 50, respectively, and overall WD10 ME's of 66.6° and 66.3° . However, the overall MB is acceptable. Similar to d04, WD10's error performance is also poor for coarser-resolution domains d03 and d02, since both domains also feature of complex topography. WD10's performance for d01 is stronger since it is less impacted by topography.

The variable Q2 is the one that is most susceptible to surface characterization changes because of the PBL schemes used. For Q2, NG yields an acceptable overall MB; however, the number of acceptable days was only 17 out of 50. DF's overall MB for Q2 does not meet the benchmark and had even fewer individual acceptable days, with 13 out of 50. [Figure 9](#) shows the Q2 and PBL diurnal profiles before and after surface characterization data improvements. In [Fig. 9a](#), the dashed lines demonstrate that prior to improving the surface characterization datasets by using ASTER and HWSD, the DF configuration and to a lesser extent the NG configuration overpredict Q2 in the morning, including the crucial hours of 0600–1000 LT. During these hours the PBL is low, but emissions are high, leading to high pollutant concentrations. In the afternoon the DF configuration nearly captures an afternoon hump feature and then overpredicts later in the day. The NG

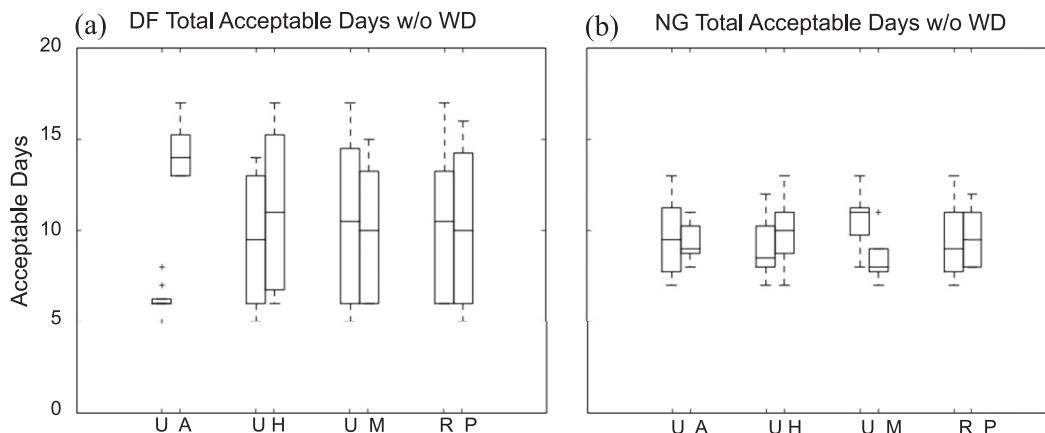


FIG. 10. Boxplot summary of (a) DF and (b) NG surface characterization changes without considering WD10 ME for terrain (USGS, U; ASTER, A), soil class (U; HWSO, H) land use (U; MODIS, M), and soil property re-initialization (R) or propagation (P). Each box represents the eight DF simulations, including their respective surface characterizations.

simulation underpredicts in the afternoon. Since these simulations use the USGS GTOPO30 terrain, which has the artificial bowl, it is expected that there would be too much moisture at the monitor locations as a result of water tending to accumulate on the western side of the city inaccurately. Observational data have shown that moisture actually accumulates more on the eastern side of the city, closer to the mountains (SDA 2013b).

When there is excess moisture, energy is lost to latent heat and therefore less energy is available for vertical mixing, leading to lower PBL heights (Pleim and Xiu 1995; Sogachev et al. 2002; Findell and Eltahir 2003). However, if the PBL scheme tends to produce an overly deep PBL, the moisture content can be artificially lowered, which happens for NG. The fact that NG underpredicts Q2 speaks to this weakness in the PBL scheme. Analyses of PBL schemes have shown that, relative to other PBL schemes, MRF tends to overdeepen the PBL (Cohen et al. 2015; Hong et al. 2006; Mass et al. 2002). NG’s PBL daily profile is deeper than DF’s, as shown by the dashed lines in Fig. 9b. The combination of model terrain and PBL scheme leads to a lower bias for specific humidity because of the terrain but for an undesired reason. Improved terrain should improve the performance for DF, but not for NG.

b. Surface characterization impacts on sMPE

Using the NG and DF configurations, we ran all combinations of surface characterizations and saw distinct sensitivity for Q2 performance and little sensitivity for NG relative to those run with DF. Tables in appendix E summarize the effects of incrementally changing the surface characterizations on sMPE.

With more accurate terrain, DF’s Q2 sMPE performance improves. This is because DF uses the YSU PBL scheme, which has been shown to better represent the PBL height for reasons explained in section 2c. With the improvement in terrain from the ASTER data, surface moisture does not tend to accumulate and lead to excess latent heat fluxes, leaving more energy available for vertical mixing and a deeper PBL. As seen in Fig. 9b, both the DF and NG PBL heights increase with applying ASTER and HWSO. However, the NG PBL height is likely too high, while the DF PBL height is more accurate and Q2 performance is improved.

Figures 10a and 10b summarize the improvements, without WD10 error, in number of acceptable days from each surface characterization. Each box represents eight simulations with one of the surface characterizations changes. The NG configuration did not significantly change number of acceptable days out of 50 with any inputs (8–11 days without WD10). Alternatively, the addition of ASTER for the DF configuration shows strong improvement in acceptability without WD10

TABLE 3. The number of days of occurrences of each cluster for observations and each cluster for the DF configuration with ASTER and HWSO data with the statistical significance *p* value for the combined periods as a test for accuracy.

Cluster	No. of observed days	No. of modeled days
L	6	3
mL	4	6
mH	13	4
H	16	26
<i>p</i> value: 0.001 35		

TABLE A1. Physics schemes available within WRF (version 3.5.1) that are used in this study.

Parameter	Scheme used	Reference
Longwave radiation	Rapid Radiative Transfer Model (RRTM)	Mlawer et al. (1997)
Shortwave radiation	Dudhia	Dudhia (1989)
Microphysics	WRF single-moment 3-class (WSM3)	Hong et al. (2004)
	Lin	Lin et al. (1983)
	Thompson	Thompson et al. (2008)
	Morrison two moment	Morrison et al. (2009)
	Pleim–Xiu (PX)	Pleim (2006)
Surface-layer scheme	MM5 similarity	Monin and Obukhov (1954)
	Noah	Ek et al. (2003)
Land surface model	Pleim–Xiu (PX)	Pleim and Xiu (1995)
	Yonsei University (YSU)	Hong et al. (2006)
PBL	Mellor–Yamada–Janjić (MYJ)	Janjić (1994)
	Asymmetric Convective Model (ACM2)	Pleim (2007)
	Markov random field (MRF)	Hong and Pan (1998)
	Bougeault–Lacarrère (BouLac)	Bougeault and Lacarrère (1989)
Cumulus convection	Kain–Fritsch scheme (KF)	Kain (2004)
	Grell cumulus scheme	Grell and Dévényi (2002)
	Betts–Miller–Janjić (BMJ)	Janjić (1994)

error as seen in Fig. 10a, caused by the significant improvement in Q2 acceptability. Following suit, the addition of HWSO leads to some more improvement. However, adding MODIS land use and soil propagation does not significantly change performance for DF or NG.

Despite the improvements gained for specific humidity performance, wind direction performance remains unacceptable for the majority of days with only 4 out of 50 days having acceptable wind direction performance for the sMPE.

5. Feature-based model performance evaluation (fMPE)

Although the model does not meet sMPE thresholds for wind direction on a majority of simulated days, the model is more likely to reproduce important meteorological features for air quality on any given day and

therefore more frequently pass an fMPE. Table 3 shows the model's ability to reproduce the frequency and accuracy of the clusters. Based on data availability requirements for clustering the observations (see methods section 2a), the February period has 15 days and October has 24 days for a total of 39 days available for analysis. The number of occurrences of each type of cluster is shown in the table for both February and October for the DF configuration with improved surface characterization data. The H cluster occurs more frequently in the model (26 days) than the observations (16 days). In addition, the number of occurrences of the mH cluster is underpredicted by the model with 4 days, as compared with the 13 days observed. This happens because mH days tend to be classified as H days and indicates that the model is predicting more occurrences of high pollution meteorology days characterized by strong synoptically driven westerly winds, and underpredicting weaker thermally

TABLE A2. Physics configurations.

	Longwave/shortwave radiation	Microphysics	Surface-layer scheme	Land surface scheme	PBL	Cumulus convection	Reference
DF	RRTM/Dudhis	WSM3	MM5	Noah	YSU	KF	Skamarock et al. (2008)
AC	RRTM/Dudhia	Lin	MM5	Noah	MYJ	KF	Kumar and Rojas (2015)
NG	RRTM/Dudhia	Thompson	MM5	Noah	MRF	Grell	Nielsen-Gammon (2001)
GP	RRTM/Dudhia	Thompson	PX	PX	ACM2	KF	Gilliam and Pleim (2010)
BL	RRTM/Dudhia	Morrison 2-m	MM5	Noah	BouLac	BMJ	Reboredo et al. (2015)

TABLE B1. Summary of physics parameterization performance metrics. Benchmarks for each metric are in parentheses. The overall values for the combined periods are shown. For individual variables, “accept of 50” is the number of days that each metric for that variable met the benchmark. The target for each number is to meet the benchmark in parentheses for each respective column.

	T2				Q2				WS				WD		
	Without Accept of 50	ME (<3.5)	MB (<2.0)	IOA (>0.8)	Accept of 50	ME (<2.0)	MB (<1.0)	IOA (>0.6)	Accept of 50	RMSE (<2.5)	MB (<1.5)	IOA (>0.6)	Accept of 50	MB (<10)	ME (<55)
DF	1	1.6	-0.2	0.9	49	1.3	1.1	0.7	13	1.5	0.2	0.6	29	-1	66.6
NG	2	1.7	-0.2	0.9	45	1.3	0.2	0.7	17	1.6	0.3	0.6	28	-2.8	66.3
BL	1	1.6	-0.2	0.9	48	1.3	1.1	0.7	14	1.7	0	0.6	22	4	68.8
AC	2	1.6	0.1	0.9	48	1.2	1.1	0.7	15	2.1	0.9	0.6	19	0.1	66.9
GP	0	2.1	-0.4	0.9	43	1.4	0.9	0.7	12	1.7	0.5	0.6	25	1.9	66.2

driven westerly winds. However, the combined number of days between the mH and H clusters is very similar between the observations and the simulation implying that the air quality model will have a similar number of above average pollution days as the observations. Also, L days are mainly misrepresented as mL days, and the observations and model show 9 and 10 days, respectively, of combined mL and L days. The fact that the model can produce a similar distribution of the clusters is a positive trait for air quality modeling purposes and may be more important than having low bias or error.

The accuracy of cluster assignments on a given day is found to be statistically significant ($p < 0.05$) using a binomial test for the combined periods. The p value for the simulation is 0.00135, and rejects the null hypothesis that every day, each cluster has an equal chance of occurrence. Furthermore, the model is able to accurately predict the cluster assignment for 15 out of the 39 days available for analysis, about 38.4% of the days. This is a much higher success rate than found with the sMPE that showed the model could only pass an sMPE 4 days out 50 (4%). The fMPE therefore demonstrates that the model can reproduce important features for air quality much more frequently than it can pass an sMPE, yielding significantly more days that can be used for air quality modeling.

6. Conclusions

In this study, we used a cluster analysis to identify meteorological features that were related to air quality in Bogotá. The driving variables for air quality were wind speed and direction and the cluster analysis showed that there were four clusters that have distinct air quality features associated with them. On low pollution days, winds were generally strongly flowing from the mountainous east leading to flushing of pollutants. Medium levels of pollution tend to occur on days when winds are thermally driven and are weaker. Pollution is higher on days when thermal winds drive transport downslope from the east in the morning and upslope from the west in the afternoon. On the highest pollution days, winds tend to be synoptically driven and flowing from the west toward the mountains. This transports pollutants to the mountains and leads to trapping as demonstrated by a case study.

Two methods were then used to evaluate WRF for air quality modeling in Bogotá, Colombia with differing results. An sMPE using community based benchmarks for T2, WS, WD, and Q2 performance showed that WRF could accurately simulate T2, but performance was mediocre for WS and Q2 and poor for WD for two

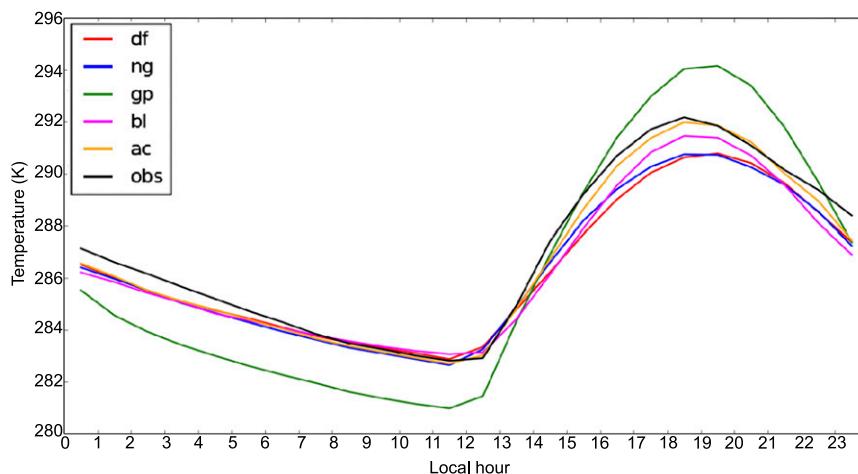


FIG. B1. Daily average temperature profile at monitor locations for the combined periods in local time. Each physics configuration is shown in a different color, with the observations in black.

types of physics configurations, DF and NG. Improvements were then made to surface characterization datasets for terrain, soil classification, and land use. With the YSU PBL scheme used by DF, Q2 performance was able to dramatically improve. The MRF PBL scheme used by the NG configuration led to an overly deep PBL relative to the DF configuration. At first a deeper PBL was favorable for Q2 performance because the terrain and soil classification datasets led to inaccurate soil moisture accumulation. The overaccumulation of soil moisture leads to lower PBLs and overpredictions of Q2. With the more accurate ASTER terrain dataset, and HWSO soil classification, excess moisture accumulation is alleviated and both configurations PBLs were raised. This led to a better representation of Q2 by the DF configuration, but not the NG configuration. With the improved surface representation the DF configuration achieved much higher performance on the sMPE for Q2. However, WD error performance was not improved with surface characterization data. In fact on the majority of days, WRF fails the sMPE for wind direction, suggesting that the sMPE does not demonstrate if WRF could reproduce the features important for air quality found by the cluster analysis.

We then ran the cluster analysis on the simulation using the DF configuration with updated surface characterization and compared the predicted clusters with the observed clusters for an fMPE. The fMPE showed that WRF could reproduce features associated with air quality at a similar frequency and distribution to the observations. WRF demonstrated a similar number of

days for the high and low clusters associated with air quality. A binomial test revealed that the accuracy was statistically significant for the simulation with a p value lower than 0.05. The fMPE showed that there were significantly more days (38.4%) when WRF could accurately simulate the features important for air quality than could pass an sMPE (<10%).

We therefore demonstrated that two evaluation techniques lead to competing conclusions on WRF's applicability to regulatory air quality modeling. WRF was able to pass an sMPE for most variables but frequently failed for specific humidity and systematically failed for wind direction. An fMPE, however, revealed that WRF could reproduce important features of wind patterns that are associated with air quality conditions. Since sMPEs often fail in regions of complex topography, the fMPE can be another way to demonstrate a model's suitability for air quality modeling. Future work will be needed to evaluate the predictive capacity of sMPE and fMPE for air quality performance. If fMPE is a better predictor of air quality modeling performance, it may be used as an alternative to sMPE. This will be most useful in settings where traditional sMPE benchmarks are not available or produce misleading results.

Acknowledgments. This project was funded by the Secretaría de Ambiente de Bogotá under Contract 1467 de 2013 with Universidad de La Salle. The viewpoint of the authors does not necessarily reflect the official thoughts of the SDA. We thank all of the people and organizations that provided information or advice for this project.

TABLE C1. Summary of physics parameterization performance metrics for 17 Feb–2 Mar across d01. Benchmarks for each metric are in parentheses. The overall values are shown. For individual variables, “accept of 15” is the number of days that each metric for that variable met the benchmark. The target for each number is to meet the benchmark in parentheses for each respective column.

Accept of 15	Without WD	T2			Q2			WS			WD					
		ME (<3.5)	MB (<2.0)	IOA (>0.8)	Accept of 15	ME (<2.0)	MB (<1.0)	IOA (>0.6)	Accept of 15	RMSE (<2.5)	MB (<1.5)	IOA (>0.6)	Accept of 15	MB (<10)	ME (<55)	Accept of 15
DF	5	2.2	-1.2	0.92	15	1.7	-0.2	0.82	15	2.7	0.3	0.76	5	-1.8	39.1	15
NG	5	2.2	-1.1	0.92	15	1.8	-0.7	0.81	15	2.7	0.3	0.76	5	0.1	39.1	15
BL	5	2.1	-1.0	0.92	15	2.0	0.3	0.79	10	2.7	0.2	0.76	4	-2.3	40.2	15
AC	4	2.2	-1.2	0.92	15	1.9	0.2	0.81	14	2.8	0.7	0.74	4	-2	39.3	15
GP	5	2.2	-1.1	0.92	15	1.7	-0.3	0.82	15	2.6	0.3	0.77	5	-2.2	39.2	15

TABLE C2. As in Table C1, but across d02.

Accept of 15	Without WD	T2			Q2			WS			WD					
		ME (<3.5)	MB (<2.0)	IOA (>0.8)	Accept of 15	ME (<2.0)	MB (<1.0)	IOA (>0.6)	Accept of 15	ME (<2.5)	MB (<1.5)	IOA (>0.6)	Accept of 15	MB (<10)	ME (<55)	Accept of 15
DF	1	1.8	-0.8	0.96	15	1.7	-1.0	0.87	14	1.6	-0.3	0.56	7	-1.4	63.8	2
NG	0	1.8	-0.8	0.96	15	1.9	-1.5	0.85	9	1.7	-0.2	0.59	8	-0.2	64.5	0
BL	0	1.8	-0.9	0.96	15	1.6	-0.1	0.89	15	1.7	-0.3	0.55	6	2.4	68	0
AC	0	1.8	-0.5	0.96	15	1.7	-0.8	0.87	14	1.9	0.3	0.53	4	1.2	64.6	2
GP	0	2.2	-0.8	0.95	15	1.5	-0.6	0.91	15	1.7	0.0	0.56	7	-4.0	64	0

TABLE C3. As in Table C1, but across d03.

	T2			Q2			WS			WD							
	Accept of 15	Without WD	ME (<3.5)	MB (<2.0)	IOA (>0.8)	Accept of 15	ME (<2.0)	MB (<1.0)	IOA (>0.6)	Accept of 15	ME (<2.5)	MB (<1.5)	IOA (>0.6)	Accept of 15	ME (<55)	MB (<10)	Accept of 15
DF	0	6	1.6	-0.2	0.95	15	1.4	-0.1	0.84	15	1.6	0	0.57	6	69.6	2.7	0
NG	0	8	1.7	-0.2	0.95	15	1.6	-0.9	0.8	15	1.6	0.2	0.6	8	68.1	4.3	0
BL	0	6	1.7	-0.2	0.95	15	1.7	0.7	0.82	12	1.7	-0.1	0.54	7	72.3	10.2	0
AC	1	6	1.7	0.1	0.95	15	1.4	0.1	0.85	15	2.0	0.7	0.54	6	67.5	6.5	1
GP	0	11	2.3	-0.2	0.93	14	1.4	0.0	0.89	15	1.7	0.3	0.58	11	67.1	-0.4	0

APPENDIX A

All Physics Configurations Used

Optimal physics configurations vary by region, period, and resolution and it is common practice to test multiple configurations (Borge et al. 2008; Ruiz et al. 2010; Carvalho et al. 2012; Evans et al. 2012). The configurations are referred to by the following acronyms: DF, AC, NG, GP, and BL. The DF configuration employs the default WRF options. Each of the other configurations was chosen based on previous studies. The AC configuration was chosen based on performance around Bogotá (Kumar and Rojas 2015). The NG configuration was chosen based on performance in a coastal environment, the Houston–Galveston region (Nielsen-Gammon 2001). The GP configuration was selected based on model skill over a range of continental conditions, including complex topography (Gilliam and Pleim 2010). The BL configuration was selected based on skill in Bogotá (Reboredo et al. 2015) and uses the Bougeault–Lacarrère (BL) PBL scheme. The full details of each physics option and configuration are summarized in Table A1 and Table A2. These five configurations were evaluated with default WRF inputs for the d04 domain.

APPENDIX B

Full Model Performance Evaluation of Physics Options

The five physics configurations show common performance for most evaluation criteria (e.g., RMSE, MB, ME for each meteorological variable). Table B1 shows that all configurations performed well for T2, moderately for WS10 and Q2, and very poorly for WD10. Figure B1 shows an average daily temperature profile for the configurations and RMCAB observations. All configurations can capture Bogotá's diurnal temperature pattern. However, GP overpredicts T2 in the afternoon and underpredicts it in the morning, which is likely caused by the propagation of soil moisture used by GP, which can lead to the misallocation of soil moisture because of poorly represented terrain. Table B1 shows that there is little variation in T2 performance between configurations. The BL and AC configurations meet all of the benchmarks on 48 out of 50 days. DF meets all benchmarks on 49 out of 50 days. NG and GP meet the benchmark on 45 and 43 days, respectively, out of 50. Also, all configurations yield acceptable IOA and ME for Q2. Only NG and GP have unacceptable MB results for Q2. Wind speed

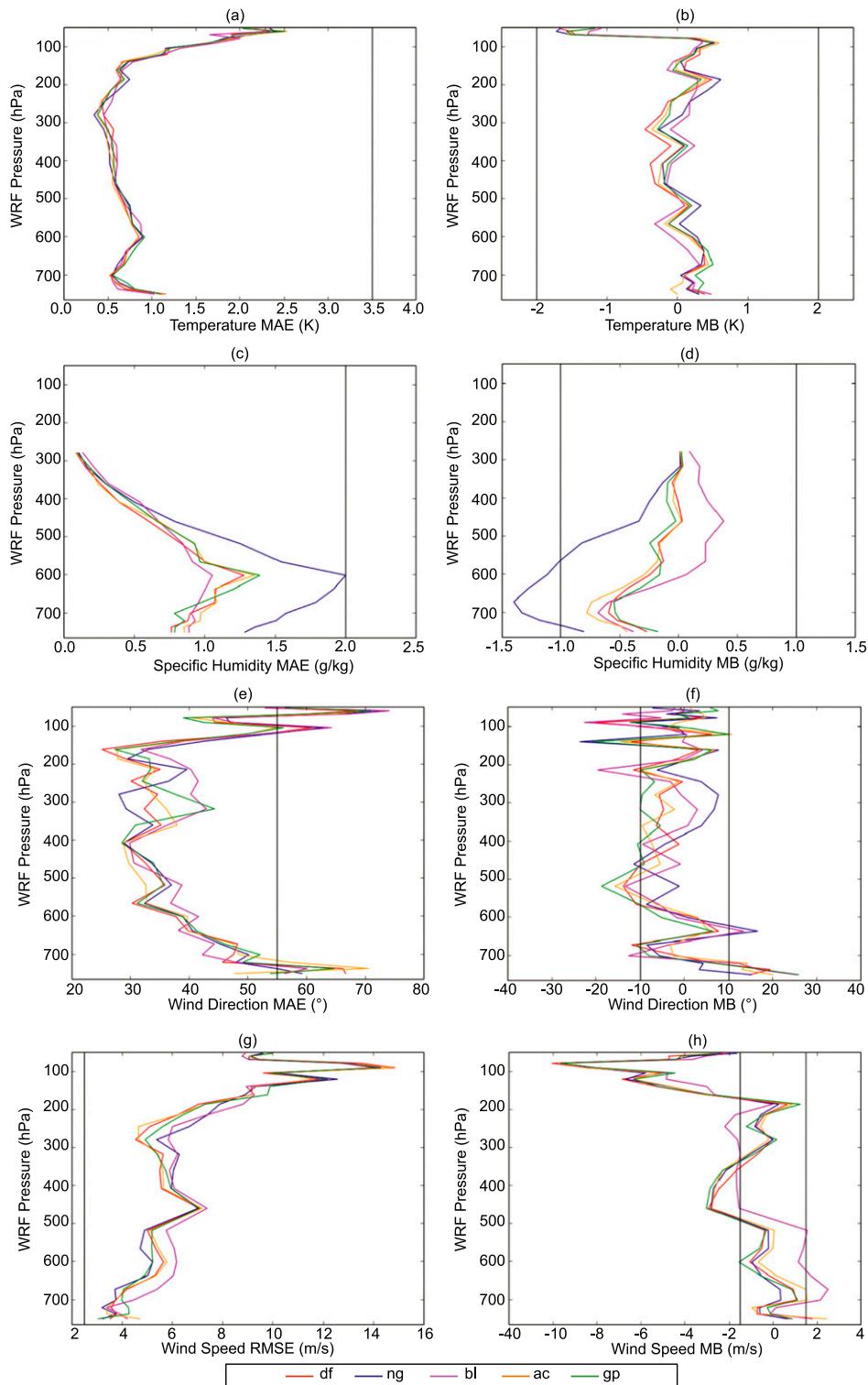


Fig. D1. Vertical profiles of temperature (a) MAE and (b) MB, specific humidity (c) MAE and (d) MB, wind direction (e) MAE and (f) MB, and wind speed (g) RMSE and (h) MB. Black vertical lines represent the desired benchmarks for model performance. The MAE and RMSE profiles would ideally be to the left of the black lines, and the MB profiles would be between them.

TABLE E1. Performance of DF simulations with varying terrain (GTOPO30, G; ASTER, A), soil class (U; HWSD, H), land use (U; MODIS, M), and soil property reinitialization (R) or propagation (P).

Key	DF		T2				O2				WS				WD		
	Accept of 50	Without WD	ME (<3.5)	MB (<2.0)	IOA (>0.8)	Accept of 50	ME (<2.0)	MB (<1.0)	IOA (>0.6)	Accept of 50	RMSE (<2.5)	MB (<1.5)	IOA (>0.6)	Accept of 50	MB (<10)	ME (<55)	Accept of 50
GUUR	1	4	1.6	-0.2	0.9	49	1.3	1.1	0.7	13	1.5	0.2	0.6	29	-1.0	66.6	4
GUUP	1	4	1.6	-0.1	0.9	49	1.3	1.0	0.7	13	1.5	0.2	0.6	29	1.7	66.1	4
GUMR	2	6	1.6	0.6	0.9	49	1.2	0.8	0.7	14	1.5	0.2	0.6	31	3.1	66.7	3
GUMP	2	6	1.7	0.7	0.9	47	1.1	0.8	0.7	15	1.5	0.2	0.6	31	3.2	66.7	3
GHUR	2	7	1.6	-0.3	0.9	48	1.3	1.0	0.7	14	1.5	0.2	0.6	30	-0.6	66.5	5
GHUP	2	7	1.6	-0.2	0.9	48	1.3	1.0	0.7	15	1.5	0.2	0.6	30	-0.5	66.7	5
GHRM	2	5	1.6	0.5	0.9	48	1.2	0.9	0.7	13	1.5	0.2	0.6	29	2.8	67	5
GHMP	2	5	1.7	0.7	0.9	47	1.2	0.8	0.7	16	1.5	0.2	0.6	29	4.0	66.8	5
AUUR	1	14	1.7	-0.7	0.9	44	1.0	0.5	0.8	25	1.5	0.1	0.7	34	3.2	66.7	4
AUUP	1	14	1.6	-0.6	0.9	45	1.0	0.5	0.8	25	1.5	0.1	0.7	34	2.5	66.5	5
AUMR	1	14	1.4	0	0.9	49	1.0	0.4	0.8	25	1.4	0.1	0.7	30	6.0	65.3	5
AUMP	1	14	1.4	0.3	0.9	49	1.0	0.4	0.8	25	1.4	0.1	0.7	29	4.9	64.3	5
AHUR	1	18	1.7	-0.8	0.9	46	1.0	0.5	0.8	26	1.4	0.1	0.7	37	2.4	65.7	4
AHUP	1	18	1.7	-0.7	0.9	46	1.0	0.5	0.8	27	1.4	0.1	0.7	30	2.3	65.6	4
AHMR	1	15	1.4	0	0.9	50	1.0	0.4	0.8	25	1.4	0.1	0.7	31	5.1	64.8	3
AHMP	1	15	1.4	0.2	0.9	49	1.0	0.4	0.8	28	1.4	0.1	0.7	27	5.7	64.7	3

performance is decent, with all configurations meeting the IOA and MB benchmarks. Last, all configurations have five or fewer acceptable days for WD. Considering all variables, the configurations only have between zero and two acceptable days.

APPENDIX C

Outer Domain Analysis

Tables C1–C3 show the outer domain model performance evaluation for d01, d02, and d03, respectively, for each physics configuration for 15 representative days in February chosen based on air quality. There is a noticeable degradation in WD10 performance from d01 to d02, and then consistently poor performance for d02, d03, and d04. The reason for this is the sudden presence of complex topography. As shown in Fig. 1, much of d02 is mountainous, and d03 is completely in the mountains, while complex topography does not cover the majority d01.

APPENDIX D

Vertical Analysis

Figure D1 shows vertical performance for WRF physics configurations for temperature MAE and MB, specific humidity MAE and MB, wind direction MAE and MB, and wind speed RMSE and MAE. Figures D1a, b show that the configurations have similar temperature error and bias profiles, and all meet the criteria throughout. For specific humidity, the NG configuration shows a lower bias than do the other configurations and a worse error. Wind direction MB is generally within the criteria throughout the vertical, with worse performance near the surface as seen using the ground monitors. Wind speed bias and RMSE tend to be worse at higher altitudes, as the absolute wind speeds increase substantially. The similarities in vertical performance suggest that surface-layer performance is most important for model differences and air quality.

APPENDIX E

Full Analysis of Surface Characterization Inputs

We show the complete analysis of the surface characterization simulations. Tables E1 and E2 show the 50-day statistics for the combined periods for all surface characterizations for the DF and NG configurations, respectively.

TABLE E2. As in Table E2, but for NG simulations.

Key	NG			T2			O2			WS			WD		
	Accept of 50	Without WD	ME (<3.5)	MB (<2.0)	IOA (>0.8)	Accept of 50	ME (<2.0)	MB (<1.0)	IOA (>0.6)	Accept of 50	RMSE (<2.5)	MB (<1.5)	IOA (>0.6)	Accept of 50	ME (<55)
GUUR	2	11	1.7	-0.2	0.9	45	1.3	0.2	0.7	17	1.6	0.3	0.6	28	-2.8
GUUP	1	10	1.7	0.0	0.9	43	1.3	0.2	0.7	19	1.6	0.4	0.6	29	-3.0
GUMR	0	9	1.7	0.6	0.9	44	1.3	0.1	0.7	17	1.5	0.2	0.6	29	0.1
GUMP	0	8	1.8	0.8	0.9	42	1.3	0.0	0.7	15	1.5	0.3	0.6	28	0.2
GHUR	2	11	1.7	-0.2	0.9	45	1.3	0.2	0.7	19	1.6	0.3	0.7	30	-1.8
GHUP	2	11	1.7	-0.1	0.9	44	1.3	0.2	0.7	18	1.6	0.4	0.6	28	-0.6
GHRM	0	8	1.7	0.6	0.9	44	1.3	0.1	0.7	14	1.5	0.3	0.6	31	0.7
GHMP	0	11	1.8	0.8	0.9	41	1.3	0.1	0.7	16	1.5	0.3	0.6	26	1.4
AUUR	2	8	1.7	-0.5	0.9	42	1.4	-0.4	0.7	14	1.5	0.3	0.7	31	0.6
AUUP	2	10	1.7	-0.4	0.9	41	1.3	-0.4	0.7	16	1.5	0.3	0.7	31	0.9
AUMR	0	9	1.5	0.2	0.9	44	1.4	-0.5	0.6	12	1.5	0.2	0.7	31	2.8
AUMP	1	10	1.5	0.5	0.9	43	1.4	-0.5	0.6	14	1.5	0.3	0.7	27	3.8
AHUR	1	9	1.7	-0.6	0.9	43	1.3	-0.3	0.7	15	1.5	0.3	0.7	33	1.2
AHUP	5	12	1.7	-0.5	0.9	44	1.4	-0.3	0.7	17	1.5	0.3	0.7	34	1.9
AHMR	0	8	1.5	0.2	0.9	46	1.4	-0.4	0.7	16	1.5	0.3	0.6	31	4.1
AHMP	1	11	1.5	0.4	0.9	43	1.4	-0.4	0.7	14	1.5	0.3	0.7	30	3.7

REFERENCES

Abrams, M., 2000: The Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER): Data products for the high spatial resolution imager on NASA's *Terra* platform. *Int. J. Remote Sens.*, **21**, 847–859, doi:10.1080/014311600210326.

Arciniegas, A., C. Rodriguez, J. E. Pachon, H. Sarmiento, and L. J. Hernandez, 2006: Estudio de la morbilidad en niños menores a 5 años por enfermedad respiratoria aguda y su relación con la concentración de partículas en una zona industrial de la ciudad de Bogotá (Study of morbidity in children younger than 5 years for acute respiratory disease and its relation to concentration of particles in an industrial area of the city of Bogotá). *Acta Nova*, **3**, 147–154.

Austin, E., A. Zanobetti, B. Coull, J. Schwartz, D. R. Gold, and P. Koutrakis, 2015: Ozone trends and their relationship to characteristic weather patterns. *J. Exposure Sci. Environ. Epidemiol.*, **25**, 532–542, doi:10.1038/jes.2014.45.

Bailey, D. T., 2000: Meteorological monitoring guidance for regulatory modeling applications. Environmental Protection Agency Rep. EPA-454/R-99-005, 168 pp. [available at <http://www.epa.gov/scram001/guidance/met/mmgrma.pdf>.]

Batjes, N. H., 2009: Harmonized soil profile data for applications at global and continental scales: Updates to the WISE database. *Soil Use Manage.*, **25**, 124–127, doi:10.1111/j.1475-2743.2009.00202.x.

Bell, M. L., L. A. Cifuentes, D. L. Davis, E. Cushing, A. G. Telles, and N. Gouveia, 2011: Environmental health indicators and a case study of air pollution in Latin American cities. *Environ. Res.*, **111**, 57–66, doi:10.1016/j.envres.2010.10.005.

Borge, R., V. Alexandrov, J. J. Del Vas, J. Lumberras, and E. Rodríguez, 2008: A comprehensive sensitivity analysis of the WRF model for air quality applications over the Iberian Peninsula. *Atmos. Environ.*, **42**, 8560–8574, doi:10.1016/j.atmosenv.2008.08.032.

Boston University, 2012: User guide for the MODIS land cover type product. Land Cover and Surface Climate Group Doc., 5 pp. [Available online at http://www.bu.edu/lcsc/files/2012/08/MCD12Q1_user_guide.pdf.]

Bougeault, P., and P. Lacarrère, 1989: Parameterization of orography-induced turbulence in a mesobeta-scale model. *Mon. Wea. Rev.*, **117**, 1872–1890, doi:10.1175/1520-0493(1989)117<1872:POOITI>2.0.CO;2.

Carvalho, D., A. Rocha, M. Gómez-Gesteira, and C. Santos, 2012: A sensitivity study of the WRF Model in wind simulation for an area of high wind energy. *Environ. Model. Software*, **33**, 23–34, doi:10.1016/j.envsoft.2012.01.019.

Chang, M., S. Fan, Q. Fan, W. Chen, Y. Zhang, Y. Wang, and X. Wang, 2014: Impact of refined land surface properties on the simulation of a heavy convective rainfall process in the Pearl River delta region, China. *Asia-Pac. J. Atmos. Sci.*, **50**, 645–655, doi:10.1007/s13143-014-0052-3.

Cheng, W. Y., and W. J. Steenburgh, 2005: Evaluation of surface sensible weather forecasts by the WRF and the Eta Models over the western United States. *Wea. Forecasting*, **20**, 812–821, doi:10.1175/WAF885.1.

Cohen, A. E., S. M. Cavallo, M. C. Coniglio, and H. E. Brooks, 2015: A review of planetary boundary layer parameterization schemes and their sensitivity in simulating southeastern U.S. cold season severe weather environments. *Wea. Forecasting*, **30**, 591–612, doi:10.1175/WAF-D-14-00105.1.

- Davis, J. M., B. K. Eder, D. Nychka, and Q. Yang, 1998: Modeling the effects of meteorology on ozone in Houston using cluster analysis and generalized additive models. *Atmos. Environ.*, **32**, 2505–2520, doi:10.1016/S1352-2310(98)00008-9.
- Dockery, D. W., C. A. Pope, X. Xu, J. D. Spengler, J. H. Ware, M. E. Fay, B. G. Ferris Jr., and F. E. Speizer, 1993: An association between air pollution and mortality in six US cities. *N. Engl. J. Med.*, **329**, 1753–1759, doi:10.1056/NEJM199312093292401.
- Dormann, C. F., and Coauthors, 2013: Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, **36**, 27–46, doi:10.1111/j.1600-0587.2012.07348.x.
- Dudhia, J., 1989: Numerical study of convection observed during the Winter Monsoon Experiment using a mesoscale two-dimensional model. *J. Atmos. Sci.*, **46**, 3077–3107, doi:10.1175/1520-0469(1989)046<3077:NSOCOD>2.0.CO;2.
- Ek, M. B., K. E. Mitchell, Y. Lin, E. Rogers, P. Grunmann, V. Koren, G. Gayno, and J. D. Tarpley, 2003: Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta Model. *J. Geophys. Res.*, **108**, 8851, doi:10.1029/2002JD003296.
- Emery, C., E. Tai, and G. Yarwood, 2001: Enhanced meteorological modeling and performance evaluation for two Texas ozone episodes. Environ Final Rep. 31984-11, 235 pp. [Available online at <https://www.tceq.state.tx.us/assets/public/implementation/air/am/contracts/reports/mm/EnhancedMetModelingAndPerformanceEvaluation.pdf>.]
- Evans, J. P., M. Ekstrom, and F. Ji, 2012: Evaluating the performance of a WRF physics ensemble over south-east Australia. *Climate Dyn.*, **39**, 1241–1258, doi:10.1007/s00382-011-1244-5.
- Findell, K. L., and E. Eltahir, 2003: Atmospheric controls on soil moisture–boundary layer interactions. Part I: Framework development. *J. Hydrometeorol.*, **4**, 552–569, doi:10.1175/1525-7541(2003)004<0552:ACOSML>2.0.CO;2.
- Gebhart, K. A., W. C. Malm, M. A. Rodriguez, M. G. Barna, B. A. Schichtel, K. B. Benedict, J. L. Collett, and C. M. Carrico, 2014: Meteorological and back trajectory modeling for the Rocky Mountain Atmospheric Nitrogen and Sulfur Study II. *Adv. Meteorol.*, **2014**, 414015, doi:10.1155/2014/414015.
- Gilliam, R. C., and J. E. Pleim, 2010: Performance assessment of new land surface and planetary boundary layer physics in the WRF-ARW. *J. Appl. Meteor. Climatol.*, **49**, 760–774, doi:10.1175/2009JAMC2126.1.
- Grell, G. A., and D. Dévényi, 2002: A generalized approach to parameterizing convection combining ensemble and data assimilation techniques. *Geophys. Res. Lett.*, **29**, doi:10.1029/2002GL015311.
- Hong, S.-Y., and H.-L. Pan, 1998: Convective trigger function for a mass-flux cumulus parameterization scheme. *Mon. Wea. Rev.*, **126**, 2599–2620, doi:10.1175/1520-0493(1998)126<2599:CTFFAM>2.0.CO;2.
- , J. Dudhia, and S. H. Chen, 2004: A revised approach to ice microphysical processes for the bulk parameterization of clouds and precipitation. *Mon. Wea. Rev.*, **132**, 103–120, doi:10.1175/1520-0493(2004)132<0103:ARATIM>2.0.CO;2.
- , Y. Noh, and J. Dudhia, 2006: A new vertical diffusion package with an explicit treatment of entrainment processes. *Mon. Wea. Rev.*, **134**, 2318–2341, doi:10.1175/MWR3199.1.
- Huang, L.-S., and R. L. Smith, 1999: Meteorologically-dependent trends in urban ozone. *Environmetrics*, **10**, 103–118, doi:10.1002/(SICI)1099-095X(199901/02)10:1<103::AID-ENV341>3.0.CO;2-D.
- Janjić, Z. I., 1994: The step-mountain eta coordinate model: Further developments of the convection, viscous sublayer, and turbulence closure schemes. *Mon. Wea. Rev.*, **122**, 927–945, doi:10.1175/1520-0493(1994)122<0927:TSMCEM>2.0.CO;2.
- Kain, J. S., 2004: The Kain–Fritsch convective parameterization: An update. *J. Appl. Meteor. Climatol.*, **43**, 170–181, doi:10.1175/1520-0450(2004)043<0170:TKCPAU>2.0.CO;2.
- Kumar, A., and N. Rojas, 2015: Statistical downscaling of WRF-Chem model: An air quality analysis over Bogota, Colombia. *Geophysical Research Abstracts*, Vol. 17, Abstract EGU2015-13722. [Available online at <http://meetingorganizer.copernicus.org/EGU2015/EGU2015-13722-1.pdf>.]
- Lin, Y. L., R. D. Farley, and H. D. Orville, 1983: Bulk parameterization of the snow field in a cloud model. *J. Appl. Meteor. Climatol.*, **22**, 1065–1092, doi:10.1175/1520-0450(1983)022<1065:BPOTSF>2.0.CO;2.
- Lu, R., and R. P. Turco, 1995: Air pollutant transport in a coastal environment—II. Three-dimensional simulations over Los Angeles basin. *Atmos. Environ.*, **29**, 1499–1518, doi:10.1016/1352-2310(95)00015-Q.
- Mann, H. B., and D. R. Whitney, 1947: On a test of whether one of two random variables is stochastically larger than the other. *Ann. Math. Stat.*, **18**, 50–60, doi:10.1214/aoms/1177730491.
- Mass, C. F., D. Ovens, K. Westrick, and B. A. Colle, 2002: Does increasing horizontal resolution produce more skillful forecasts? *Bull. Amer. Meteor. Soc.*, **83**, 407–430, doi:10.1175/1520-0477(2002)083<0407:DIHRPM>2.3.CO;2.
- McNally, D. E., 2009: Meteorological modeling performance guidelines to replace old standards. *Proc. 10th Annual Ad-Hoc Meteorological Modelers Meeting*, Boulder, CO, Environmental Protection Agency. [Available online at <http://www.epa.gov/scram001/adhoc/mcnally2009.pdf>.]
- Mlawer, E. J., S. J. Taubman, P. D. Brown, M. J. Iacono, and S. A. Clough, 1997: Radiative transfer for inhomogeneous atmosphere: RRTM, a validated correlated-*k* model for the longwave. *J. Geophys. Res.*, **102**, 16 663–16 682, doi:10.1029/97JD00237.
- Monin, A. S., and A. Obukhov, 1954: Basic laws of turbulent mixing in the surface layer of the atmosphere. *Contrib. Geophys. Inst. Acad. Sci. USSR*, **151**, 163–187.
- Morrison, H., G. Thompson, and V. Tatarskii, 2009: Impact of cloud microphysics on the development of trailing stratiform precipitation in a simulated squall line: Comparison of one and two-moment schemes. *Mon. Wea. Rev.*, **137**, 991–1007, doi:10.1175/2008MWR2556.1.
- NCEP, 2000: NCEP FNL Operational Model Global Tropospheric Analyses, continuing from July 1999. National Center for Atmospheric Research Computational and Information Systems Laboratory Research Data Archive, accessed 7 February 2014, doi:10.5065/D6M043C6.
- , 2014: Meteorological Assimilation Data Ingest System. NCEP Central Operations, subset used: February 2012–October 2012, accessed March 2014. [Available online at https://madis.noaa.gov/madis_api.shtml.]
- Nielsen-Gammon, J. W., 2001: Initial modeling of the August 2000 Houston–Galveston ozone episode. Report to the Technical Analysis Division, Texas Natural Resource Conservation Commission 19, 71 pp. [Available online at https://www.drivecleanmachine.org/assets/public/implementation/air/am/contracts/reports/mm/Init_MM5_Modeling_2000Aug.pdf.]
- Nikolakopoulos, K. G., D. A. Vaipoulos, and G. A. Skianis, 2004: ASTER DTM vs GTOPO30 and 1/250,000 topographic

- maps—The case of Lefkas Island, Greece. *Proc. Geoscience and Remote Sensing Symp.*, Anchorage, AK, Institute of Electrical and Electronics Engineers, doi:10.1109/IGARSS.2004.1370807.
- Pachón, J. E., H. Sarmiento, and T. Hoshiko, 2013: Health risk represented by inhaling polycyclic aromatic hydrocarbons (PAH) during daily commuting involving using a high traffic flow route in Bogota. *Salud Pública*, **15**, 398–407.
- Pleim, J. E., 2006: A simple, efficient solution of flux-profile relationships in the atmospheric surface layer. *J. Appl. Meteor. Climatol.*, **45**, 341–347, doi:10.1175/JAM2339.1.
- , 2007: A combined local and nonlocal closure model for the atmospheric boundary layer. Part I: Model description and testing. *J. Appl. Meteor. Climatol.*, **46**, 1383–1395, doi:10.1175/JAM2539.1.
- , and A. Xiu, 1995: Development and testing of a surface flux and planetary boundary layer model for application in meso-scale models. *J. Appl. Meteor. Climatol.*, **34**, 16–32, doi:10.1175/1520-0450-34.1.16.
- Reboredo, B., R. Arasa, and B. Codina, 2015: Evaluating sensitivity to different options and parameterizations of a coupled air quality modelling system over Bogotá, Colombia. Part I: WRF Model configuration. *Open J. Air Pollut.*, **4**, 47–64, doi:10.4236/ojap.2015.42006.
- Rincon, A., and N. Rojas, 2014: Exploratory application of CCATT-BRAMS modeling system for the metropolitan area of Bogotá: Preliminary representation of meteorology and transport. Abstracts, *IV Congreso Colombiano y Conferencia Internacional de Calidad del Aire y Salud Pública*, Bogotá, Colombia, CASAP Technical Committee, 323–324. [Available online at http://casap.com.co/2013/en/docs/programa_final.pdf.]
- , and —, 2015: BRAMS model for a tropical inter-andean region. Presentation slides, *V Congreso Colombiano y Conferencia Internacional de Calidad del Aire y Salud Pública*, Bucaramanga, Colombia, CASAP Technical Committee. [Available online at http://casap.com.co/documentos/Memorias_13_agosto/440PM_Salon5_RinconA.pdf.]
- Rodríguez, C., A. Arciniegas, J. Pachón, H. Sarmiento, and L. J. Hernández, 2006: Relationship between acute respiratory illness and air pollution in Bogotá. *Revista Investigaciones En Seguridad Social Y Salud*, **8**.
- Ruiz, J. J., C. Saulo, and J. Noguez-Paegle, 2010: WRF model sensitivity to choice of parameterization over South America: Validation against surface variables. *Mon. Wea. Rev.*, **138**, 3342–3355, doi:10.1175/2010MWR3358.1.
- SDA, 2010: Plan decenal de descontaminación del aire para Bogotá. Anexo del Decreto “Por medio del cual se adopta el Plan Decenal de Descontaminación del Aire para Bogotá” (Ten-year air decontamination plan for Bogotá. Annex to the Decree “The means by which the Ten-Year Plan of Air Decontamination for Bogotá is adopted”). Secretaria de Ambiente, Bogota, Colombia, 94 pp. [Available online at http://www.alcaldiabogota.gov.co/sisjur/adminverblobawa?tabla=T_NORMA_ARCHIVO&p_NORMFIL_ID=930&f_NORMFIL_FILE=X&inputfileext=NORMFIL_FILENAME.]
- , 2013a: Red de Monitoreo de Calidad del Aire de Bogotá (Air quality monitoring network of Bogotá). Secretaria Distrital de Ambiente de Bogota, accessed 14 February 2014. [Available online at <http://201.245.192.252:81>.]
- , 2013b: 2012 Annual Air Quality Network report. Secretaria Distrital de Ambiente de Bogota Rep. 126PM04-PR84-M-A2-V1.0, 176 pp.
- Skamarock, W. C., and Coauthors, 2008: A description of the Advanced Research WRF version 3. NCAR Tech. Note NCAR/TN-475+STR, 113 pp., doi:10.5065/D68S4MVH.
- Sogachev, A., G. V. Menzhulin, M. Heimann, and J. O. N. Lloyd, 2002: A simple three-dimensional canopy–planetary boundary layer simulation model for scalar concentrations and fluxes. *Tellus*, **54B**, 784–819, doi:10.1034/j.1600-0889.2002.201353.x.
- Teixeira, J. C., A. C. Carvalho, T. Luna, M. J. Carvalho, and A. Rocha, 2014: Sensitivity of the WRF Model to the lower boundary in an extreme precipitation event—Madeira island case study. *Nat. Hazards Earth Syst. Sci.*, **14**, 2009–2025, doi:10.5194/nhess-14-2009-2014.
- Tesche, T. W., D. E. McNally, C. A. Emery, and E. Tai, 2001: Evaluation of the MM5 model over the midwestern U.S. for three 8-hour oxidant episodes. Report prepared for Kansas City Ozone Technical Workgroup by Alpine Geophysics LLC and ENVIRON International Corp., 151 pp.
- Thompson, G., P. R. Field, R. M. Rasmussen, and W. D. Hall, 2008: Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part II: Implementation of a new snow parameterization. *Mon. Wea. Rev.*, **136**, 5095–5115, doi:10.1175/2008MWR2387.1.
- Thompson, M. L., J. Reynolds, L. H. Cox, P. Guttorp, and P. D. Sampson, 2001: A review of statistical methods for the meteorological adjustment of tropospheric ozone. *Atmos. Environ.*, **35**, 617–630, doi:10.1016/S1352-2310(00)00261-2.
- Tibshirani, R., G. Walther, and T. Hastie, 2001: Estimating the number of clusters in a data set via the gap statistic. *J. Roy. Stat. Soc.*, **63B**, 411–423, doi:10.1111/1467-9868.00293.
- WHO, 2006: WHO quality guidelines for particulate matter, ozone, nitrogen dioxide and sulfur dioxide: Global update 2005. World Health Organization, 20 pp. [Available online at http://apps.who.int/iris/bitstream/10665/69477/1/WHO_SDE_PHE_OEH_06.02_eng.pdf.]
- Yao, T., J. H. Fung, H. Ma, A. K. H. Lau, P. W. Chan, J. Z. Yu, and J. Xue, 2014: Enhancement in secondary particulate matter production due to mountain trapping. *Atmos. Res.*, **147–148**, 227–236, doi:10.1016/j.atmosres.2014.05.007.
- Yucel, I., 2006: Effects of implementing MODIS land cover and albedo in MM5 at two contrasting U.S. regions. *J. Hydrometeorol.*, **7**, 1043–1060, doi:10.1175/JHM536.1.